Classification et extension automatique d'annotations d'images en utilisant un réseau Bayésien

Sabine Barrat, Salvatore Tabbone

Classification et extension automatique d'annotations d'images en utilisant un réseau Bayésien

Sabine Barrat[1] – Salvatore Tabbone[1]

LORIA-UMR7503 - Université Nancy 2
BP 239 - 54506 Vandœuvre-les-Nancy Cedex, France

{barrat,tabbone}@loria.fr

Résumé : *Dans nombre de problèmes de vision, au lieu d'avoir des données d'apprentissage entièrement annotées, il est plus facile d'obtenir seulement un sous-ensemble de données dotées d'annotations, car ceci est moins restrictif pour l'utilisateur. Pour ces raisons, dans ce papier, nous considérons le problème de classification d'images faiblement annotées, où seulement un petit sous-ensemble de la base de données est annoté par des mots-clés. Nous présentons et évaluons une nouvelle méthode qui améliore l'efficacité de la classification d'images par le contenu, en intégrant des concepts sémantiques extraits du texte, et en étendant automatiquement les annotations existantes à des images non annotées ou faiblement annotées. Notre modèle s'inspire de la théorie des modèles graphiques probabilistes, permettant de traiter les données manquantes. Les résultats de la classification visuo-textuelle, obtenus grâce à une base d'images provenant d'Internet, partiellement et manuellement annotées, montrent une amélioration de* 32.3% *en terme de taux de reconnaissance, par rapport à la classification visuelle. De plus, l'extension automatique d'annotations, avec notre modèle, à des images faiblement annotées, augmente encore le taux reconnaissance de* 6.8%.

Mots-clés : modèles graphiques probabilistes, réseaux Bayésiens, sélection de variables, classification, annotation automatique

1 Introduction

La croissance rapide d'Internet et de l'information multimédia a engendré un besoin en techniques de recherche d'information multimédia, et plus particulièrement en recherche d'images. On peut distinguer deux tendances. La première, appelée recherche d'images par le texte, consiste à appliquer des techniques de recherche de textes à partir d'ensembles d'images complètement annotés. La seconde approche, appelée recherche d'images par le contenu, est un domaine plus récent et utilise une mesure de similarité (similarité de couleur, forme ou texture) entre une image requête et une image du corpus utilisé. Afin d'améliorer la reconnaissance, une solution consiste à combiner les informations visuelles et sémantiques : on parle d'approches visuo-textuelles. Plusieurs chercheurs ont déjà exploré cette possibilité : Barnard et al. [BAR 03] segmentent les images en régions. Chaque région est représentée par un ensemble de caractéristiques visuelles et un ensemble de mots-clés. Les images sont alors classifiées en modélisant de façon hiérarchique les distributions de leurs mots-clés et caractéristiques visuelles. Grosky et al. [GRO 01] associent des coefficients aux mots afin de réduire la dimensionnalité. Les vecteurs de caractéristiques visuelles et les vecteurs d'indices correspondant aux mots-clés sont concaténés pour procéder à la recherche d'images. Benitez et al. [BEN 02] extraient de la connaissance à partir de collections d'images annotées, en classifiant les images représentées par leurs caractéristiques visuelles et textuelles. Des relations entre les informations visuelles et textuelles sont alors découvertes. Enfin, l'annotation automatique d'images peut être utilisée dans les systèmes de recherche d'images, pour organiser et localiser les images recherchées ou pour améliorer la classification visuo-textuelle. Cette méthode peut être vue comme un type de classification multi classes avec un grand nombre de classes, aussi large que la taille du vocabulaire. Plusieurs travaux ont été proposés dans ce sens. On peut citer, sans être exhaustif, les méthodes basées sur la classification [GAO 06, YAN 06], les méthodes probabilistes [BLE 03, FEN 04] et l'affinement d'annotations [WAN 06, RUI 07]. Dans cette direction, la contribution de ce papier est de proposer une méthode de classification d'images, en utilisant une approche visuo-textuelle et en étendant automatiquement des annotations existantes à des images faiblement annotées. L'approche proposée est dérivée de la théorie des modèles graphiques probabilistes. Nous introduisons une méthode pour traiter le problème des données manquantes dans le contexte d'images annotées par mots-clés comme défini en [BLE 03, KHE 04]. L'incertitude autour de l'association entre un ensemble de mots-clés et une image est représentée par une distribution de probabilité jointe sur le vocabulaire et les caractéristiques visuelles extraites de notre collection d'images couleurs ou à niveaux de gris. Or les réseaux Bayésiens sont un moyen simple de représenter une distribution de probabilité jointe d'un ensemble de variables aléatoires, de visualiser les propriétés de dépendance conditionnelle, et ils permettent d'effectuer des calculs complexes comme l'apprentissage des probabilités et l'inférence, avec des manipulations

graphiques. Un réseau Bayésien semble donc approprié pour représenter et classifier des images associées à des mots-clés. Enfin, étant donnée la taille des caractéristiques visuelles, l'utilisation d'un algorithme de sélection de variables est introduite : il permet d'augmenter notre taux de reconnaissance de 4.5% en utilisant seulement les caractéristiques les plus pertinentes et en réduisant le problème de dimensionnalité.

2 Un réseau Bayésien pour la classification d'images faiblement annotées

Nous présentons un modèle hiérarchique probabiliste multimodal (images et mots-clés associés) pour classifier de grandes bases de données d'images annotées. Les caractéristiques visuelles sont considérées comme des variables continues, et les éventuels mots-clés associés comme des variables discrètes. Le modèle proposé est un modèle de mélange de lois multinomiales et de densités à mélange de Gaussiennes (notons "modèle de mélange GM-M"). En effet, l'observation de plusieurs pics sur les histogrammes de variables caractéristiques nous a conduits à considérer que les caractéristiques visuelles peuvent être estimées par des densités de type mélange de Gaussiennes. Les variables discrètes correspondant aux éventuels mots-clés sont supposées suivre une distribution multinomiale sur le vocabulaire des mots-clés.

Soit F un échantillon d'apprentissage composé de m individus $f_{1_i}, ..., f_{m_i}, \forall i \in \{1, ..., n\}$, où n est la dimension des signatures obtenues par concaténation des vecteurs caractéristiques issus du calcul des descripteurs sur chaque image de l'échantillon. Chaque individu $f_j, \forall j \in \{1, ..., m\}$ est caractérisé par n variables continues. Un contexte de classification supervisée est considéré donc les m individus sont divisés en k classes $c_1, ..., c_k$. Soient $G_1, ..., G_g$ les g groupes dont chacun a une densité Gaussienne avec une moyenne $\mu_l, \forall l \in \{1, ..., g\}$ et une matrice de covariance $\sum_l$. De plus, soient $\pi_1, ..., \pi_g$ les proportions des différents groupes, $\theta_l = (\mu_l, \sum_l)$ le paramètre de chaque Gaussienne et $\Phi = (\pi_1, \pi_1, ..., \pi_g, \theta_1, ..., \theta_g)$ le paramètre global du mélange. Alors la densité de probabilité de F conditionnellement à la classe $c_i, \forall i \in \{1, ..., k\}$ est définie par $P(f, \Phi) = \sum_{l=1}^{g} \pi_l p(f, \theta_l)$ où $p(f, \theta_l)$ est la Gaussienne multivariée définie par le paramètre θ_l.

Ainsi nous avons un modèle de mélange de Gaussiennes (GMM) par classe. Ce problème peut être représenté par le modèle probabiliste de la Figure 1, où :

- Le nœud "Classe" est un nœud discret, pouvant prendre k valeurs correspondant aux classes prédéfinies $c_1, ..., c_k$.
- Le nœud "Composante" est un nœud discret correspondant aux composantes (i.e les groupes $G_1, ..., G_g$) des mélanges. Cette variable peut prendre g valeurs, i.e le nombre de Gaussiennes utilisé pour calculer les mélanges. Il s'agit d'une variable latente qui représente le poids de chaque groupe (i.e les $\pi_l, \forall l \in \{1, ..., g\}$).
- Le nœud "Gaussienne" est une variable continue représentant chaque Gaussienne $G_l, \forall l \in \{1, ..., g\}$ avec son propre paramètre $(\theta_l = (\mu_l, \sum_l))$. Il correspond à l'ensemble des vecteurs caractéristiques dans chaque classe.
- Enfin les arêtes représentent l'effet de la classe sur le paramètre de chaque Gaussienne et son poids associé. Le cercle vert sert à montrer la relation entre le modèle graphique proposé et les GMMs : nous avons un GMM (entouré en vert), composé de Gaussiennes et de leur poids associé, par classe.

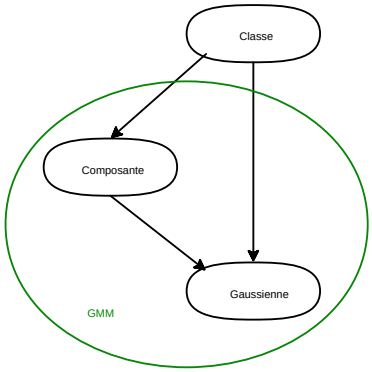

FIG. 1 – GMMs représentés par un modèle graphique probabiliste

Maintenant le modèle peut être complété par les variables discrètes, notées $KW\ 1, ..., KW\ n$, correspondant aux éventuels mots-clés associés aux images. Des *a priori* de Dirichlet [ROB 97], ont été utilisés pour l'estimation de ces variables. C'est-à-dire que l'on introduit des pseudo comptes supplémentaires à chaque instance de façon à ce qu'elles soient toutes virtuellement représentées dans l'échantillon d'apprentissage. Ainsi chaque instance, même si elle n'est pas représentée dans l'échantillon d'apprentissage, aura une probabilité non nulle. Comme les variables continues correspondant aux caractéristiques visuelles, les variables discrètes correspondant aux mots-clés sont incluses dans le réseau en les connectant à la variable classe.

Notre classificateur peut alors être décrit par la Figure 2. La variable latente "α" montre qu'un *a priori* de Dirichlet a été utilisé. La boîte englobante autour de la variable KW indique n répétitions de KW, pour chaque mot-clé.

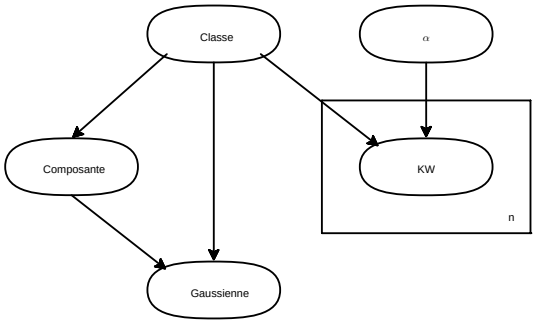

FIG. 2 – Modèle de mélange GM-M

2.1 Apprentissage des paramètres et inférence

L'algorithme EM a été utilisé pour apprendre les paramètres des mélanges de Gaussiennes. Mais le problème majeur réside dans le traitement des données manquantes. En effet, seulement certaines données sont complètement observées. C'est le cas de toutes les caractéristiques visuelles pour les images couleurs, ou des caractéristiques de forme seulement, pour les images à niveaux de gris. Par contre les caractéristiques couleur pour les images à niveaux de gris, et surtout les mots-clés pour un grand nombre d'images, sont manquants. Concernant les caractéristiques couleur, les valeurs manquantes sont clairement distribuées de façon homogène, puisqu'elles correspondent aux images à niveaux de gris. Au contraire les valeurs manquantes sont distribuées aléatoirement pour les variables $KW\ i$, $\forall i \in \{1, ..., n\}$. Ce genre de problème peut également être traité par l'algorithme EM. Le but général de cet algorithme, expliqué en détail dans [DEM 77], consiste à calculer, de manière itérative, le maximum de vraisemblance, quand les observations peuvent être vues comme des données incomplètes. Un algorithme d'inférence est également nécessaire pour classifier de nouvelles images. En effet, le processus d'inférence consiste à calculer les distributions de probabilité *a posteriori* d'un ou plusieurs sous-ensembles de nœuds. Dans le cas de la classification, la valeur du nœud classe est inférée. Conformément à la structure de notre réseau Bayésien, le processus d'inférence propage les valeurs du niveau des caractéristiques d'une image, représenté par le nœud "Gaussienne", en passant par le nœud "Composante" et les nœuds mots-clés, jusqu'au niveau du nœud "Classe". Un algorithme de passage de message [KIM 83] est appliqué au réseau. Dans cette technique, chaque noeud est associé à un processeur qui peut envoyer des messages de façon asynchrone à ses voisins jusqu'à ce qu'un équilibre soit atteint. Ainsi une image requête f_j, caractérisée par ses caractéristique visuelles $v_{j_1}, ..., v_{j_m}$ et ses éventuels mots-clés $KW\ 1_j, ..., KW\ n_j$ est considérée comme une "évidence" représentée par $P(f_j) = P(v_{j_1}, ..., v_{j_m}, KW\ 1_j, ..., KW\ n_j) = 1$ quand le réseau est évalué. Grâce à l'algorithme d'inférence, les probabilités de chaque nœud sont mises à jour en fonction de cette évidence. Après la propagation de croyances, on connaît $\forall i \in \{1, ..., k\}$, la probabilité *a posteriori* $P(c_i|f_j)$. L'image requête f_j est affectée à la classe c_i maximisant cette probabilité.

2.2 Extension automatique d'annotations d'images

Etant donnée une image sans mot-clé, ou faiblement annotée, le modèle proposé peut être utilisé pour calculer une distribution des mots-clés conditionnellement à une image et ses éventuels mots-clés existants. En effet, pour une image f_j annotée par $k, \forall k \in \{0, ..., n\}$ mots-clés, où n est le nombre maximum de mots-clés par image, l'algorithme d'inférence permet de calculer la probabilité *a posteriori* $P(KW_{i_j}|f_j, KW_1, ..., KWk), \forall i \in \{k+1, ..., n\}$. Cette distribution représente une prédiction des mots-clés manquants d'une image. Par exemple, considérons le Tableau 1 présentant 3 images avec leurs éventuels mots-clés existants et les mots-clés obtenus après l'extension automatique d'annotations. La première image, sans mot-clé, a été automatiquement annotée par deux mots-clés appropriés. De même, la seconde image, annotée au départ par deux mots-clés, a vu son annotation s'étendre à trois mots-clés. Le nouveau mot-clé, "coucher de soleil" est approprié. Enfin la troisième image, initialement annotée par un mot-clé, a obtenu deux nouveaux mots-clés grâce à l'extension automatique d'annotations. Le premier nouveau mot-clé "nuage" est correct, par contre le second, "coucher de soleil", ne convient pas. Cette erreur est probablement due au grand nombre d'images de la base annotées par les trois mots-clés "pont", "nuage" et "coucher de soleil", et à l'algorithme d'inférence.

3 Réduction de dimensionnalité

Les larges dimensions des vecteurs de caractéristiques visuelles engendrent un problème de dimensionnalité. En effet, une trop grande dimension des vecteurs caractéristiques provoque un mauvais apprentissage des mélanges de Gaussiennes, car il y a une disproportion entre la taille de l'échantillon d'apprentissage et la dimension des vecteurs. Pour résoudre ce problème, nous avons adapté une méthode de réduction de dimensionnalité, qui permet d'extraire les caractéristiques les plus pertinentes et discriminantes, avec une perte minimale d'information. La méthode de régression, appelée LASSO (Least Absolute Shrinkage and Selection Operator) [TIB 96], a été choisie pour sa stabilité et sa facilité de mise en œuvre. De plus cette méthode permet surtout de sélectionner des variables, contrairement à l'Analyse en Composantes Principales (ACP), par exemple. Le LASSO réduit les coefficients de régression en imposant une pénalité sur leur taille. Ces coefficients minimisent la somme des erreurs quadratiques avec un seuil associé à la somme des valeurs absolues des coefficients $\beta^{lasso} = \arg\min_\beta \sum_{i=1}^{N}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2$ avec la contrainte $\sum_{j=1}^{p} |\beta_j| \leq s$.

Le LASSO utilise une pénalité $L_1 : \sum_{j=1}^{p} |\beta_j|$. Cette contrainte implique que pour des petites valeurs de s, $s \geq 0$, certains coefficients β_j vont s'annuler. Ainsi choisir s est similaire à choisir le nombre de variables explicatives dans un modèle de régression. Les variables correspondant aux coefficients non nuls sont sélectionnées. Les solutions du LASSO ont été calculées avec l'algorithme "Least Angle Regression" (LAR) [EFR 04]. Cet algorithme exploite la structure particulière du LASSO, et fournit un moyen efficace de calculer simultanément les solutions pour toutes les valeurs de s. La forme linéaire du LASSO a été utilisée dans une étape de prétraitement, sur les caractéristiques visuelles, totalement indépendamment de notre classificateur Bayésien. Pour adapter cette méthode à notre problème, considérons

image	mots-clés initiaux	mots-clés après extension automatique d'annotations
		pont eau
	pont nuage	pont nuage coucher de soleil
	pont	pont nuage coucher de soleil

TAB. 1 – Exemple d'images et de leurs éventuels mots-clés, avant et après extension automatique d'annotations

nos données d'apprentissage : y_i représente la somme des caractéristiques du vecteur moyen de la classe c_i, et $x_j = \{x_{j_1}, ..., x_{j_p}\}$ les p caractéristiques de l'individu j.

4 Résultats expérimentaux

Dans cette section, nous présentons une évaluation de notre modèle sur plus de 3000 images provenant d'Internet, et aimablement fournies par Kherfi et al. [KHE 04]. Ces images sont réparties en 16 classes. Par exemple la Figure 3 présente quatre images de la classe "cheval".

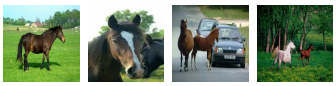

FIG. 3 – Exemples d'images de la classe "cheval"

65% de la base a été annotée manuellement par 1 mot-clé, 28% par 2 mots-clés et 6% par 3 mot-clés, en utilisant un vocabulaire de 39 mots-clés. Par exemple, parmi les quatre images de la Figure 3, la première est annotée par 2 mots-clés, "animal" et "cheval". La seconde est annotée par 1 mot-clé seulement : "animal". Les deux autres images n'ont aucune annotation. Les caractéristiques visuelles utilisées sont issues d'un descripteur de couleur, un histogramme de couleurs, et d'un descripteur de forme basé sur les transformées de Fourier/Radon. Notre méthode a été évaluée en effectuant cinq validations croisées, dont chaque proportion de l'échantillon d'apprentissage est fixée à 25%, 35%, 50%, 65% et 75% de la base. Les 75%, 65%, 50%, 35% et 25% respectivement restants sont retenus pour l'échantillon de test. Dans chaque cas les tests ont été répétés 10 fois, de façon à ce que chaque observation ait été utilisée au moins une fois pour l'apprentissage et les tests. Pour chacune des 5 tailles de l'échantillon d'apprentissage, on calcule le taux de reconnaissance en effectuant la moyenne des taux de reconnaissance obtenus pour les 10 tests. Tout d'abord la sélection de variables avec la méthode du LASSO nous a permis de réduire significativement le nombre de variables initial. Le Tableau 2 montre le nombre de variables sélectionnées pour chaque descripteur avec la méthode du LASSO, comparé à celui obtenu avec la méthode "Sequential Forward Selection" (notée SFS) [PUD 94]. Le Tableau 3 montre l'impact de la méthode de sélection de variables sur la qualité de la classification. De façon à mesurer cet impact, une classification a été effectuée sur les caractéristiques visuelles avec notre modèle (noté mélange GM-M), et deux autres classificateurs : un classificateur SVM classique [CHA 01], et un algorithme flou des k plus proches voisins (noté FKNN) [KEL 85]. Puis nous avons comparé les taux de reconnaissance pour ces 3 classificateurs sans sélection de variables préalable et après la sélection d'un sous-ensemble de variables avec les méthodes SFS et LASSO. L'algorithme flou des k plus proches voisins a été exécuté avec $k = 1$ et $k = m$, où m désigne le nombre moyen d'images par classe dans l'échantillon d'apprentissage. On peut constater que la méthode SFS a permis de sélectionner moins de variables que la méthode LASSO (voir Tableau 2). Cependant les résultats du Tableau 3 montrent que la sélection de variables avec la méthode LASSO améliore le taux de reconnaissance de 1.5% en moyenne par rapport à celui obtenu sans sélection de variables préalable, et de 7.2% en moyenne comparé à celui obtenu après sélection de variables avec la méthode SFS. De plus cette amélioration est observée quelle que soit la méthode de classification utilisée. Ainsi, seules les variables sélectionnées avec la méthode du LASSO ont été utilisées dans la suite des expérimentations. Considérons maintenant le Tableau 4. La notation "C + F" signifie que les descripteurs de forme et de couleur ("C" pour couleur et "F" pour forme) ont été combinés. La notation "C + F + KW" indique la combinaison des informations visuelles et textuelles. Les taux de reconnaissance confirment que la combinaison des caractéristiques visuelles et sémantiques est toujours plus performante que l'utilisation d'un seul type d'information. En effet, on observe que la combinaison des caractéristiques visuelles et des mots-clés (quand ils sont disponibles) augmente le taux de reconnaissance d'environ 38.5% comparé aux résultats obtenus avec le descripteur couleur seul, de 58% comparé à la classification basée sur le descripteur de forme et de 37% par rapport à la classification utilisant uniquement l'information textuelle. De plus on peut noter que pour toutes les expérimentations, combiner les deux descripteurs visuels apporte en moyenne une amélioration de 16% du taux de reconnaissance, comparé à l'utilisation d'un seul. En-

fin, la classification visuo-textuelle montre une amélioration d'environ 32.3% en terme de taux de reconnaissance, par rapport à la classification basée sur l'information visuelle seule. Ensuite, le Tableau 5 montre l'efficacité de notre approche (mélange GM-M) comparée aux classificateurs SVM et FKNN. Les résultats ont été obtenus en utilisant les deux caractéristiques visuelles et les éventuels mots-clés associés. Il apparaît que les résultats du mélange GM-M sont meilleurs que ceux du SVM et du FKNN. Enfin, des annotations ont été ajoutées automatiquement à toutes les images de la base de façon à ce que chacune soit annotée par 3 mots-clés. Puis, afin d'évaluer la qualité de cette extension d'annotations, la classification visuo-textuelle a été répétée avec les mêmes spécifications que dans le tableau 4. Le Tableau 6 montre l'efficacité de notre extension automatique d'annotations. En effet, les taux de reconnaissance après l'extension d'annotations sont toujours meilleurs qu'avant. De plus l'extension automatique d'annotations améliore le taux de reconnaissance de 6.8% en moyenne.

5 Conclusion et perspectives

Nous avons proposé un modèle efficace permettant de combiner l'information visuelle et textuelle, de traiter les données manquantes et d'étendre des annotations existantes à d'autres images. De plus nous avons adapté la méthode du LASSO, qui a résolu notre problème de dimensionnalité et ainsi diminué la complexité de la méthode. Le LASSO nous a permis d'améliorer le taux de reconnaissance, comparé à une méthode de sélection de variables plus classique. Nos expérimentations ont été effectuées sur une base d'images partiellement annotées provenant d'Internet. Les résultats montrent que la classification visuo-textuelle a amélioré le taux de reconnaissance comparée à la classification basée sur l'information visuelle seule. De plus notre réseau Bayésien a été utilisé pour étendre des annotations à d'autres images, ce qui a encore amélioré le taux de reconnaissance. Enfin la méthode proposée s'est montrée compétitive avec des classificateurs classiques. Les futurs travaux seront dédiés à la considération des préférences des utilisateurs, par la mise en place d'un processus de retour de pertinence. Plus précisément, les préférences de l'utilisateur pourraient être représentées par une mise à jour des paramètres du réseau (i.e les probabilités de chaque variable en fonction de la dernière observation classifiée), pendant le processus d'inférence.

6 Biliographie

Références

[BAR 03] Barnard K., Duygulu P., Forsyth D., De Freitas N., Blei D. M., Jordan M. I., Matching words and pictures, *Journal of Machine Learning Research*, vol. 3, n° 6, pp. 1107–1135, 2003.

[BEN 02] Benitez A., Shih-Fu C., Perceptual knowledge construction from annotated image collections, *in ICME '02*, vol. 1, pp. 189–192, 2002.

[BLE 03] Blei D. M., Jordan M. I., Modeling annotated data, *in SIGIR '03*, pp. 127–134, 2003.

[CHA 01] Chang C.-C., Lin C.-J., LIBSVM : a library for support vector machines, 2001.

[DEM 77] Dempster A. P., Laird N. M., Rubin D. B., Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, n° 1, pp. 1–38, 1977.

[EFR 04] Efron B., Hastie T., Johnstone I., Tibshirani R., Least Angle Regression, *Annals of Statitics*, vol. 32, pp. 407–499, 2004.

[FEN 04] Feng S., Manmatha R., Lavrenko V., Multiple Bernoulli relevance models for image and video annotation, *in CVPR '04*, vol. 2, pp. 1002–1009, 2004.

[GAO 06] Gao Y., Fan J., Xue X., Jain R., Automatic image annotation by incorporating feature hierarchy and boosting to scale up SVM classifiers, *in ACM MULTIMEDIA '06*, pp. 901–910, 2006.

[GRO 01] Grosky W. I., Zhao R., Negotiating the Semantic Gap : From Feature Maps to Semantic Landscapes, *in SOFSEM '01*, pp. 33–52, 2001.

[KEL 85] Keller J., Gray M., Givens J., A fuzzy k-nearest neighbor algorithm, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, n° 4, pp. 580–585, 1985.

[KHE 04] Kherfi M. L., Brahmi D., Ziou D., Combining Visual Features with Semantics for a More Effective Image Retrieval, *in ICPR '04*, vol. 2, pp. 961–964, 2004.

[KIM 83] Kim J. H., Pearl J., A computational model for combined causal and diagnostic reasoning in inference systems, *IJCAI-83*, pp. 190–193, 1983.

[PUD 94] Pudil P., Novovičová J., Kittler J., Floating search methods in feature selection, *Pattern Recognition Letters*, vol. 15, n° 11, pp. 1119–1125, 1994, Elsevier Science Inc.

[ROB 97] Robert C., *A decision-Theoretic Motivation*, Springer-Verlag, 1997.

[RUI 07] Rui X., Li M., Li Z., Ma W.-Y., Yu N., Bipartite graph reinforcement model for web image annotation, *in ACM MULTIMEDIA '07*, pp. 585–594, 2007.

[TIB 96] Tibshirani R., Regression Shrinkage and Selection Via the Lasso, *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, n° 1, pp. 267–288, 1996.

[WAN 06] Wang C., Jing F., Zhang L., Zhang H.-J., Image annotation refinement using random walk with restarts, *in ACM MULTIMEDIA '06*, pp. 647–650, 2006.

[YAN 06] Yang C., Dong M., Hua J., Region-based Image Annotation using Asymmetrical Support Vector Machine-based Multiple-Instance Learning, *in CVPR '06*, pp. 2057–2063, 2006.

Nombre de variables	Descripteur couleur	Descripteur de forme
Sans sélection	48	180
SFS	11	7
LASSO	45	23

TAB. 2 – Nombre moyen de variables en fonction de la méthode de sélection de variables

Méthode de sélection de variables	SVM	FKNN $k = 1$	FKNN $k = m$	mélange GM-M
Sans sélection	32.2	43.2	39	40.7
SFS	30.5	33.7	35	33.8
LASSO	32.6	44.1	39.3	45.22

TAB. 3 – Taux de reconnaissance moyens (en %) pour les classificateurs SVM, FKNN et le mélange GM-M, en fonction de la méthode de sélection de variables

Spécifications		Couleur	Forme	Mots-clés	C + F	C + F + KW
proportion apprentissage	proportion test					
25%	75%	35	17.8	36.6	39.4	69.7
35%	65%	36.9	18.1	38.9	42.2	74.4
50%	50%	38.7	18.5	41.1	45	79.1
65%	35%	41.1	20.6	41.5	46.6	81.7
75%	25%	43.5	21.8	45.1	52.9	82.9

TAB. 4 – Taux de reconnaissance (en %) de la classification visuelle vs. classification visuo-textuelle (avec mélange GM-M)

Spécifications		SVM	FKNN $k = 1$	FKNN $k = m$	mélange GM-M
proportion apprentissage	proportion test				
25%	75%	38.3	59.1	58.5	69.7
35%	65%	41.3	62.3	58.3	74.4
50%	50%	39.9	68.2	58.2	79.1
65%	35%	40.5	72.9	67	81.7
75%	25%	41.9	73.2	69.3	82.9

TAB. 5 – Taux de reconnaissance (en %) des classificateurs SVM et FKNN vs. notre mélange GM-M

Spécifications		Avant extension d'annotations	Après extension d'annotations
proportion apprentissage	proportion test		
25%	75%	69.7	77
35%	65%	74.4	79.3
50%	50%	79.1	85.4
65%	35%	81.7	87.6
75%	25%	82.9	92.7

TAB. 6 – Taux de reconnaissance (en %) de la classification visuo-textuelle (avec mélange GM-M) avant et après extension automatique d'annotations